Entwicklung & Optimierung von LLM-Anwendungen (Python).

Von Architekturprinzipien bis Agenten: Sie entwickeln LLM-Systeme und setzen DSGVO-konforme KI produktiv um.

Seminar	6 Termine verfügbar	Teilnahmebescheinigung
Präsenz / Virtual Classroom	24 Unterrichtseinheiten	Online durchführbar

Seminarnummer: 31381

Stand: 01.11.2026. Alle aktuellen Informationen finden Sie unter https://akademie.tuv.com/s/31381

Dieses Training vermittelt AI Engineers, Data Scientists und Softwareentwicklern die Grundlagen und Best Practices für die Entwicklung leistungsstarker LLM-Anwendungen in Python. Es vermittelt praxisnah moderne Architekturprinzipien und -muster, darunter RAG-Architekturen, Agentensysteme und fortgeschrittenes Prompt-Engineering bis hin zu Deployment und LLMOps. Dabei werden ebenso wichtige Themen wie DSGVO-Konformität, Monitoring und Responsible AI behandelt.

Die Teilnehmer lernen, wie sie moderne KI-Systeme effizient aufbauen, konfigurieren und skalieren können – von der Auswahl passender Modelle und Quantisierungstechniken bis zur Integration von Vektordatenbanken und der Orchestrierung von Multi-Agenten-Systemen. Praktische Übungen runden das Training ab, sodass ein direkter Transfer in reale Anwendungsfälle erleichtert wird.

Dieses Training bietet eine kompakte Wissensbasis, um KI-Projekte sicher und performant zu realisieren, von der Architektur über die Umsetzung bis zum Betrieb und der Governance.

Nutzen

- Sie erlernen Architekturen und Best Practices für skalierbare LLM-Anwendungen.
- Sie entwickeln RAG-Systeme mit Vektordatenbanken eigenständig.
- Sie planen und setzen sichere, DSGVO-konforme KI-Integrationen um.
- Sie erhalten eine Einführung in LLMOps für Monitoring, Fine-Tuning und CI/CD.
- Sie sammeln Erfahrung mit Multi-Agent-Systemen und praxisnahen Projekten.

Zielgruppe

AI Engineers, Data Scientists & Softwareentwickler mit Python-Erfahrung, die LLM-Anwendungen entwickeln und produktiv einsetzen wollen, wie

® TÜV, TUEV und TUV sind eingetragene Marken. Eine Nutzung und Verwendung bedarf der vorherigen Zustimmung.

- Software Engineers (Python)
- Machine Learning Engineers
- Data Scientists
- Architekten für KI-Systeme.

Voraussetzungen

Gute Python-Kenntnisse (Fortgeschrittenen-Niveau)

Abschluss

Teilnahmebescheinigung

Teilnahmebescheinigung der TÜV Rheinland Akademie

Inhalte des Seminars

Tag 1:
Grundlagen, Architektur & Prompting

- Architektur moderner LLM-Anwendungen
- Überblick: RAG, Agenten-Architekturen, Orchestrierungsmuster
- Best Practices & Design Patterns für skalierbare KI-Systeme
- Trade-offs: Latenz, Kosten, Qualität

Prompt Engineering intensiv

- Grundlagen: Few-Shot, Zero-Shot, Chain-of-Thought
- Erweiterte Techniken: ReAct, Self-Ask, Structured Outputs, Guardrails
- Hinweis auf weiterführende Techniken (Tree-of-Thought, Graph-of-Thought)
- Hands-on: Erste Prototypen mit OpenRouter, OpenWebUI& Python

Modelle, Quantisierung & Konfiguration

- Überblick Modellfamilien (LLaMA, Mistral, Phi, Falcon, Mixtral, etc.)
- Parametergrößen (7B, 13B, 70B …) und ihr Einfluss auf Qualität & Hardwarebedarf
- Quantisierung (Q4, Q8 …) für Performance & Speicheroptimierung

® TÜV, TUEV und TUV sind eingetragene Marken. Eine Nutzung und Verwendung bedarf der vorherigen Zustimmung.

- Model Configuration: Temperatur, Top-K/P, Prompt-Templates, System-Prompts

Tag 2:

RAG & Deployment

- Retrieval-Augmented Generation (RAG)
- Embeddings & Chunking-Strategien
- Vektordatenbanken: FAISS, Weaviate, Pinecone, Milvus
- Advanced RAG: Ranking, Hybrid Retrieval, Query Expansion
- Optimierung: Latency-Reduktion, Kontextfenster-Strategien, Guardrails
- Hands-on: Aufbau & Optimierung eines RAG-Systems mit LangChain

Deployment & Infrastruktur

- Lokale & Cloud-Bereitstellung: Vor- & Nachteile
-
- Deployment-Server: openLLM, vLLM, llama.cpp im Vergleich
- Containerisierung & Skalierung mit Docker/Kubernetes
- Monitoring & Observability: GPU-Auslastung, Kosten, Antwortqualität

Tag 3:

LLMOps, Agenten & Praxisprojekt

- LLMOps & Governance
- Monitoring & Evaluation: automatisierte Benchmarks, Halluzinationserkennung
- Fine-Tuning & Parameter-Efficient Tuning (LoRA/QLoRA)
- CI/CD & Testing von LLM-Anwendungen
- DSGVO-Compliance, Responsible AI & Audit-Fähigkeit

Multi-Agent-Systeme & MCP

- Monitoring & Evaluation: automatisierte Benchmarks, Halluzinationserkennung
- Fine-Tuning & Parameter-Efficient Tuning (LoRA/QLoRA)
- CI/CD & Testing von LLM-Anwendungen
- DSGVO-Compliance, Responsible AI & Audit-Fähigkeit

Wichtige Hinweise

® TÜV, TUEV und TUV sind eingetragene Marken. Eine Nutzung und Verwendung bedarf der vorherigen Zustimmung.

Die Seminarunterlagen werden in englischer Sprache bereitgestellt. Die Kurssprache ist Deutsch.

Terminübersicht und Buchung

Buchen Sie Ihren Wunschtermin jetzt direkt online unter https://akademie.tuv.com/s/31381 und profitieren Sie von diesen Vorteilen:

- Schneller Buchungsvorgang
- Persönliches Kundenkonto
- Gleichzeitige Buchung für mehrere Teilnehmer:innen

Alternativ können Sie das Bestellformular verwenden, um via Fax oder E-Mail zu bestellen.

® TÜV, TUEV und TUV sind eingetragene Marken. Eine Nutzung und Verwendung bedarf der vorherigen Zustimmung.